

Introducing Software Heritage, the Library of Alexandria for Code

 slate.com/blogs/future_tense/2016/06/30/software_heritage_from_inria_wants_to_preserve_old_versions_of_computer.html

By Jacob Brogan

[Slate](#)

The Citizen's Guide to the Future

June 30 2016 3:29 PM

future  tense

62



Can we stop old code from fading away?

Software Heritage/Inria

Like so many other elements of our digital culture, software tends to disappear rapidly, new versions supplanting old ones, even as unused and obsolete programs slip into disrepair. Where the [Internet Archive's Wayback Machine](#) has long collected past versions of websites, there's never been a similar repository for code. The new initiative [Software Heritage](#) is attempting to change that, pushing back against what it describes as the fundamental fragility of software. The site, which promotes itself as a centralized database for source code, already archives more than 2 billion source files, drawn from millions of projects.

As Software Heritage [explains](#) on its website, it aspires "to collect, preserve, and share all software that is publicly available in source code form." Drawn as it is primarily from a variety of existing repositories such as GitHub, this isn't some-one stop shop for obsolete software. You can't, for example, just drop in to download old versions of MS Word or Final Cut, however much you might prefer them to the programs' latest instantiations. In that regard, it's unlike, say, [Abandonia](#), which collects ancient DOS games whose creators no longer support them. Instead, Software Heritage will primarily serve as a resource for coders, one that will help them maintain a sense of their discipline's constantly developing history.

ADVERTISING

[inRead](#) invented by [Teads](#)

Software Heritage is a project of [Inria](#), the French national institute for computer science and applied mathematics, but a number of other organizations have already [expressed support](#), including Microsoft and the

Linux Foundation. Many of those companies and institutions have issued statements that resonate with Inria's central premises, with Microsoft, for one, claiming that it believes the project "will help curate and conserve human knowledge in the form of code for future generations as well as help today's generations of developers find and re-use code worldwide."

Though it presents itself as the software Library of Alexandria, for now, at least, the archive is relatively daunting, especially for the non-coder: It encourages visitors to check whether their own work is already in the database, but you can't casually download programs, or even easily discern what it contains if you don't know what you're looking for. Moving ahead, however, it intends to organize things in a way that makes them more accessible—and in a way that emphasizes especially important projects and resources. It invites visitors to help it expand its coverage and laying out a variety of other features that it plans to develop, including full-text search.

Among other ideals, Software Heritage operates on the premise that preserving older software is important for the sciences. Since software is often important to replicating previous experiments, preserving it serves as a crucial means of pushing back against the [reproducibility crisis](#). As the organization [explains](#) on its website, "Software Heritage will ensure availability and traceability of software, the missing vertex in the triangle of scientific preservation," since it will help future researchers to know and employ "the exact version of the software used" by their predecessors.

In other words, Software Heritage is a project that has real, practical potential, offering an important reminder that we shouldn't take the ephemeral quality of the internet for granted.